BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

The 2.8 Å crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: a five-layered αβαβα structure constructed from two core domains of identical topology

M Mathieu[1], JPh Zeelen[1], RA Pauptit[1†], R Erdmann[2‡], W-H Kunau[2] and RK Wierenga[1*]

[1]EMBL, Postfach 10.2209, D-69012 Heidelberg, Germany and [2]Institut für Physiologische Chemie, Abteilung Zellbiochemie, Ruhr Universität Bochum, D-44780 Bochum, Germany

Background: The peroxisomal enzyme 3-ketoacyl-coenzyme A thiolase of the yeast *Saccharomyces cerevisiae* is a homodimer with 417 residues per subunit. It is synthesized in the cytosol and subsequently imported into the peroxisome where it catalyzes the last step of the β-oxidation pathway. We have determined the structure of this thiolase in order to study the reaction mechanism, quaternary associations and intracellular targeting of thiolases generally, and to understand the structural basis of genetic disorders associated with human thiolases.

Results: Here we report the crystal structure of unliganded yeast thiolase refined at 2.8 Å resolution. The enzyme comprises three domains; two compact core domains having the same fold and a loop domain. Each of the two core domains is folded into a mixed five-stranded β-sheet covered on each side by helices and the two are assembled into a five-layered αβαβα structure. The central layer is formed by two helices, which point with their amino termini towards the active site. The loop domain, which is to some extent stabilized by interactions with the other subunit, runs over the surface of the two core domains, encircling the active site of its own subunit.

Conclusions: The crystal structure of thiolase shows that the active site is a shallow pocket, shaped by highly conserved residues. Two conserved cysteines and a histidine at the floor of this pocket probably play key roles in the reaction mechanism. The two active sites are on the same face of the dimer, far from the amino and carboxyl termini of both subunits and the disordered amino-terminal import signal sequence.

Structure 15 September 1994, 2:797–808

Key words: coenzyme A, inborn errors, peroxisome, targeting sequence, thiolase

Introduction

Thiolases are widely distributed in nature, being found in both prokaryotes and eukaryotes [1,2]. In the latter, they occur in three different compartments; the cytosol, the mitochondria and the peroxisomes. There are two classes of thiolase; those involved in biodegradation [3-ketoacyl-coenzyme A (CoA) thiolase or thiolase I (EC 2.3.1.16), which catalyzes the last reaction in the β-oxidation pathway], and those invloved in biosynthetic pathways [acetoacetyl-CoA thiolase or thiolase II (EC 2.3.1.9)] (see Fig. 1). The biosynthetic and biodegradative enzymes exhibit sequence similarities and key active site residues are conserved, suggesting that they have similar reaction mechanisms, but kinetic differences have been observed [3]. However, there may be substantial differences between the binding pockets, because 3-ketoacyl-CoA thiolases have relatively large substrates consisting of a long 3-keto fatty acid molecule covalently linked to the CoA moiety, while the substrate for the acetoacetyl-CoA thiolases is smaller because the bulky fatty acid is replaced by the acetoacetyl moiety.

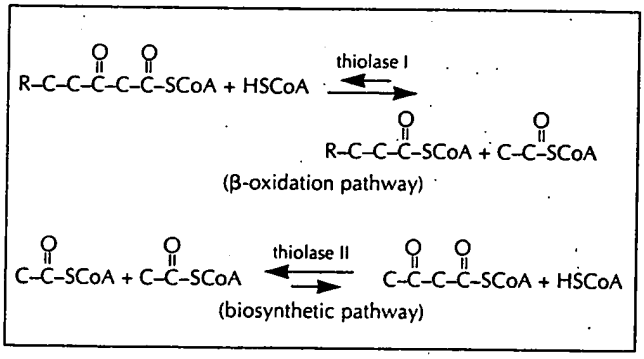

Fig. 1. The reactions catalyzed by thiolase I (3-ketoacyl-CoA thiolase) and thiolase II (acetoacetyl-CoA thiolase). The cleavage reaction is thermodynamically favored in both cases. The reaction catalyzed by thiolase II is a carbon–carbon bond forming reaction, known as a Claisen condensation [45].

There are several reasons for our interest in the structure of thiolases. Firstly, the details of the reaction mechanism are not fully understood. Early investigations on

*Corresponding author. Present addresses: †Zeneca Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK and ‡Laboratory of Cell Biology, Rockefeller University, New York, USA.

 ISSN 0969-2126

mitochondrial pig heart thiolase [redacted] and thiolase I [5–7] suggested the involvement of essential and reactive sulfhydryl groups in the reaction mechanism and led to the proposal of a two-step mechanism with an acyl-S enzyme as an obligatory intermediate. A cysteine residue of pig heart acetoacetyl-CoA thiolase was the first to be identified as the possible candidate to carry this acyl group [4]. This cysteine residue, Cys125 in *Saccharomyces cerevisiae* thiolase (Fig. 2), corresponds to a cysteine which is strictly conserved in all known thiolase sequences. Subsequent studies suggested that at least two sulfhydryl groups may actually be present near the active site with only one of them being involved in the formation of an acyl–enzyme intermediate during catalysis [6]. From spectroscopic measurements with thiolase I from porcine heart it was concluded that the distance between the two active site cysteines should be smaller than 14 Å [8]. The importance of cysteine residues has been confirmed by recent investigations on the bacterial thiolase II of *Zoogloea ramigera* which showed that two cysteines and a histidine are important for catalysis [9].

Secondly, thiolases have different quaternary structures. The reasons for this are completely unknown. Because thiolases occur in prokaryotes as well as in three cellular compartments in eukaryo[redacted]nalysis of the available sequences has been used to study evolutionary relationships [10]. The peroxisomal 3-ketoacyl-CoA thiolases always occur as dimers, but in the thiolase family as a whole, dimers, tetramers and hexamers [11] are observed. For example, the mitochondrial porcine thiolase I [5] and thiolase II [12] are both homotetramers and prokaryotic thiolase is part of a heterotetrameric β-oxidation complex, α2β2, in which a multi-functional polypeptide (the α-subunit) is associated with thiolase I (the β-subunit) [2].

Thirdly, the intracellular targeting signals of peroxisomal and mitochondrial thiolases investigated to date differ from those of the majority of the matrix proteins of their respective organelles. Mitochondrial thiolase I is synthesized with a non-cleavable amino-terminal import signal. Most of the known peroxisomal matrix proteins contain their targeting information within their three carboxy-terminal amino acids (peroxisomal targeting signal 1, PTS1). In contrast, peroxisomal thiolases of rat liver [13,14] and *S. cerevisiae* [15,16] have been shown to possess an amino-terminal targeting signal (PTS2). The amino-terminal peptide of the rat enzyme is cleaved off after import into the peroxisome, but this has not yet been unequivocally demonstrated for the yeast thiolase.

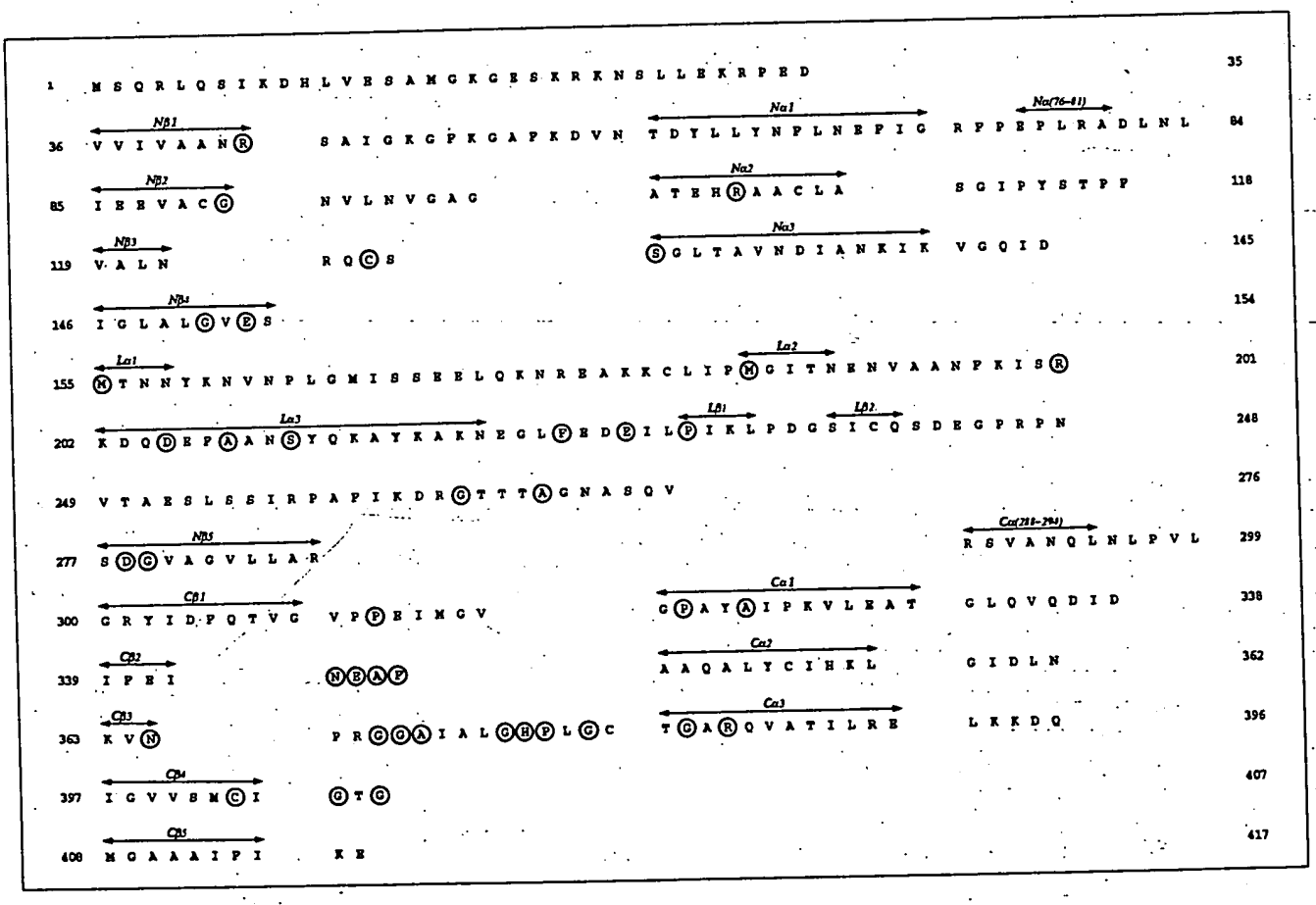

Fig. 2. The sequence of *S. cerevisiae* thiolase [10]. The β-strands and α-helices are numbered as in Fig. 4, with the prefix N; L or C indicating the N-domain, loop domain and C-domain, respectively. The residues conserved in 21 thiolase sequences are circled. The layout of the figure has been chosen to highlight the regular structural features of the two core domains.

Here we report the 2.8 Å crystal structure of the dimeric peroxisomal thiolase I of *S. cerevisiae* (hereafter referred to as yeast thiolase), a thiolase that has a broad chain-length specificity. The protein fold is described in detail and compared with folds of other proteins of known structure. The implications of the structure for biochemical studies on protein targeting, for the causes of genetic diseases due to defects in the thiolase gene, and for the reaction mechanism are discussed.

Results

The asymmetric unit of crystallized thiolase contains one yeast thiolase dimer. The current structure has been built in an averaged 3.1 Å multiple isomorphous replacement including anomalous scattering (MIRAS) map and refined at 2.8 Å resolution. The refinement statistics are shown in Table 1. The overall shape of the dimer is shown schematically in Fig. 3. The two subunits are tightly associated by interactions across the dimer two-fold axis. The two active sites are on one side of the molecule and the amino and carboxyl termini are on the opposite side. The largest cross section of the molecule is observed at the active site face, which has dimensions of approximately 80 Å × 55 Å. The thickness, measured perpendicular to the active site face, is ~50 Å (Fig. 3). Three domains can be recognized in each subunit — two core domains and a loop domain. The electron density map clearly defines the complete chain tracing of the core domains, but only an incomplete trace can be deduced for the loop domain.

Table 1. Refinement statistics.

Protein atoms	5201
Solvent atoms	0
Resolution range	8 Å–2.8 Å
R-factor[a] (no. of reflections)	19.8 % (16 479)
Free R-factor[b] (no. of reflections)	33.4 % (1792)
Geometric parameters	
Rms deviation from ideal bond lengths	0.012 Å
Rms deviation from ideal bond angles	2.04°
$\chi_1-\chi_2$ imperfection values[c]	
subunit-1	34.4°
subunit-2	34.5°
Residues with ϕ,ψ outside allowed region	11
Rms ΔB for all atoms	5.0 Å²
backbone atoms	3.5 Å²
side-chain atoms	6.9 Å²
Average B-factor all atoms	25.1 Å²
main-chain atoms	23.6 Å²
side-chain atoms	29.3 Å

[a] $R_{factor} = \Sigma|F_{OBS}-F_{CALC}|/\Sigma|F_{OBS}|$. [b] The free R-factor has been calculated from 1792 reflections which were not included in the X-PLOR calculations throughout the refinement. [c] The $\chi_1-\chi_2$ imperfection value [44] is the rms difference between observed $\chi_1-\chi_2$ values and the nearest preferred cluster values, as observed in a database of well refined structures.

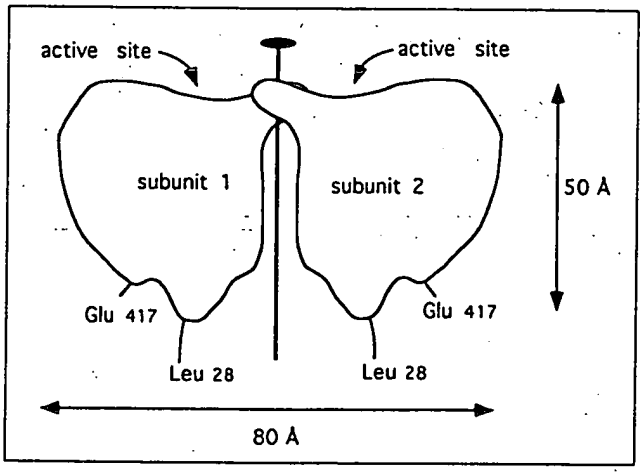

Fig. 3. Schematic diagram of the thiolase dimer. In this side view, the dimer two-fold axis runs vertically, between the two subunits, the amino terminus (Leu28) and the carboxyl terminus (Glu417) are at the bottom and the active sites of both subunits are at the top.

The thiolase fold

The three domains of the thiolase subunit are built up from four different sequence fragments (Fig. 4). The residues at the amino terminus are mobile. The first residue of the model of both subunits is Leu28. Residues Leu28 to Ser154 fold into the amino-terminal core domain (the N-domain). Subsequent residues, Met155 to Val276, are folded into a long loop (the loop domain) with some α-helical and β-strand regions. Residues Ser277 to Arg287 complete the N-domain. Residues Arg288 to Glu417 fold into the carboxy-terminal core domain (the C-domain). According to this domain subdivision, the N-domain, the loop domain, and the C-domain consist of 138, 122 and 130 residues, respectively.

The two core domains have the same topology — a five-stranded mixed β-sheet covered by helices on both sides (Fig. 4). This topology is actually very simple with the following secondary structure elements observed sequentially: βαβαβαββ, referred to as β1α1β2α2β3α3β4β5 (Fig. 2). The five β-strands form a mixed β-sheet consisting of one antiparallel and four parallel β-strands. The topology can be written as +3x,+1x,−2x,−1, according to the nomenclature of Richardson [17]. The three crossover connections are three α-helices; α1 and α2 on one side of the sheet and α3 on the other. In the C-domain the turn between β-strand 4 and β-strand 5 (abbreviated to Cβ4 and Cβ5) is just a short turn, consisting of three residues. However, in the N-domain, this is the place where the loop domain is inserted. In the N-domain, β3 is well defined and is the edge strand of the sheet in the subunit, but in the dimer it is also hydrogen bonded to its two-fold related equivalent strand of the second subunit. In the C-domain, β3 is located at the surface with just a few hydrogen bonds between it and Cβ2.

The N-domain and C-domain can be superimposed on each other via an approximate two-fold axis (by rotating through 178° and translating along the two-fold axis by

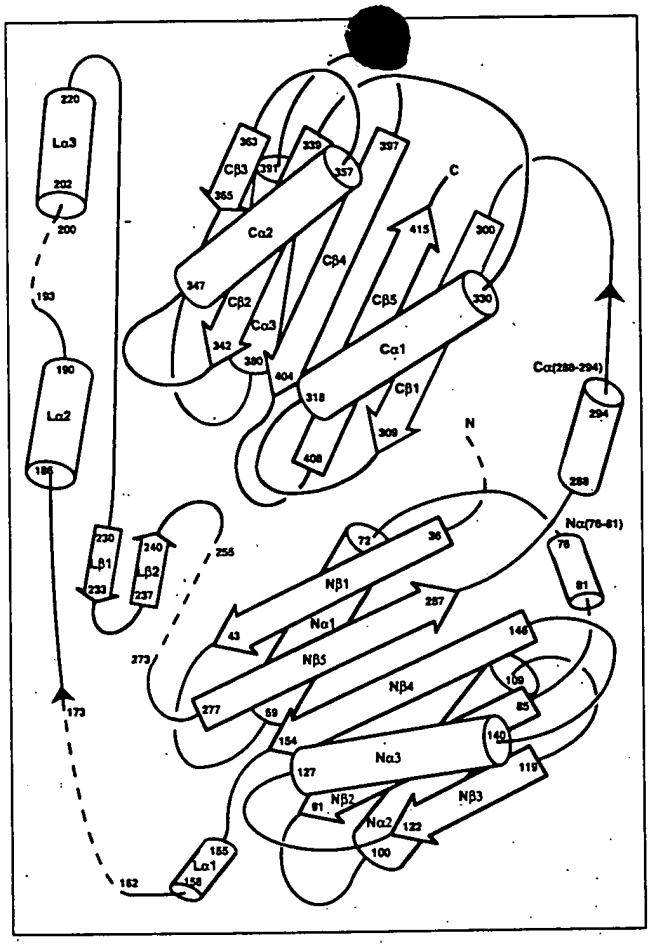

Fig. 4. The secondary structure elements. The broken lines indicate the regions which could not be built owing to lack of density. The nomenclature of the secondary structure elements is the same as in Fig. 2.

1.0 Å) (Fig. 5). Most of the equivalent residues are in the β-strands and α-helices of the two core domains. The root mean square (rms) fit for the 73 equivalenced Cα atoms is 1.7 Å although the sequence identity is only 11 %. The domain two-fold axis runs approximately parallel to helix α3. Assembly of the N- and C-domains generates a compactly folded, five-layered structure with the α3 helices in the center layer, sandwiched between the two mixed sheets (Fig. 6). Helices Nα1 and Nα2 and Cα1 and Cα2 form the two outer layers.

The three α-helical layers each contain two helices. All six helices point in the same direction. The two α3 helices of the central layer are orientated with their amino termini towards the active site. The loop domain is also on this side of the molecule (Fig. 6).

The electron density map does not provide a complete chain tracing for the loop domain. In both subunits, residues 162–173 at the beginning of this loop (immediately after Nβ4), residues 193–200 (in the middle of the domain) and residues 255–273 at the end of the loop domain (just before Nβ5) are disordered (see Fig. 4). Some other loop domain residues in subunit-1 (248–254) and subunit-2 (174,184,185) also could not be built. The structure of the ordered part of the loop domain is not the same in both subunits (as shown in Fig. 7), which correlates with differences in crystal contacts of the two subunits. The loop domain residues have higher mean B-factors than other regions of the structure (Fig. 8). The loop domain encircles the active site (Fig. 9). A substantial fraction of the loop domain residues could not be built (40 out of 122), but some secondary structure is still present (Fig. 4). After the second break in the main chain, the longest helix (Lα3; residues 202–220) of the entire molecule is seen. This helix lies at one extremity of the molecule, far from the dimer two-fold axis (Fig. 9). The beginning and end of the loop domain are close to the active site pocket (Figs 9 and 10).

Active site pocket

The position of the active site pocket can be located from the patterns of conserved residues (Fig. 2) and also from studies of the reaction mechanism of the homologous *Z. ramigera* thiolase, which have identified two cysteines and a histidine as being important for catalysis [9]. The equivalent residues in yeast thiolase are Cys125, His375 and Cys403. These residues form the base of a pocket surrounded by the loop domain. Most of the active site residues are located in the C-domain (Table 2).

Fig. 5. Superposition of the N-domain (green) and C-domain (red). The numbers refer to the N-domain, except for 375 and 403 which indicate the positions of the catalytic residues His375 and Cys403 in the C-domain. The N-domain (residues 28 to 287) is interrupted between residues 154 and 277 by the presence of the loop domain.

A detailed view of this pocket is shown in Fig. 10. Cys403 is rather buried, whereas Cys125 and His375 are more exposed to solvent. At a resolution of 2.8 Å not all the atomic positions can be known accurately; therefore the distances given below are approximate, but they do provide important information about the active site architecture, as visualized in Fig. 10. The side chains of Cys125 and His375 are close together; for example, in subunit-1 the distance between Sγ(Cys125) and Nε2(His375) is 4.1 Å. Other polar atoms near Sγ(Cys125) are O(Cys403) and N(Gly405) which are 4.4 Å and 3.6 Å away, respectively. The distance from Nε2(His375) to Sγ(Cys403) is 7.5 Å and the distance between Sγ(Cys125) and Sγ(Cys403) is 6.2 Å. Polar atoms close to Sγ(Cys403) are O(Met315), N(Gly316), O(Gly316), Oε1(Gln349) and Nε2(Gln349) which are 4.1 Å, 4.2 Å, 4.0 Å, 4.2 Å and 3.9 Å away, respectively. As shown in Fig. 10, the side chain of Cys403 is between the side chains of Met315 and Gln349. Interestingly, Met315 and Gln349 are completely conserved in 16 (out of 21) thiolase sequences and are replaced by phenylalanine and valine respectively in the other five known sequences.

The floor of the active site pocket is defined by the residues His375, Cys125 and Cys403. The center of the active site is chosen as the center of mass of Sγ(Cys125), Sγ(Cys403) and Nε2(His375). At least two more methionines are close to the active site; Met402 (just below the floor of the active site) and Met155 (near His375). Met408 is also near the active site. The Sδ atoms of all five methionines of the ordered part of thiolase actually occur within 11 Å of the active site center. In addition, five (out of eight) Sγ atoms of cysteines are within 13 Å of the active site. The relevance of this preferred occurrence of methionines and cysteines near the active site is not clear.

There are no charged polar atoms of lysine, arginine, glutamate or aspartate residues within 10 Å of the active site. The nearest atom of this kind is Oε1(Glu153), which is 12 Å away. The Glu153 side chain is actually shielded from the active site pocket by Leu377. Solvent accessibility studies show that Glu153 has no solvent accessible surface. Its Oε1 and Oε2 atoms are hydrogen bonded to N(Val93) and N(Leu94), and to N(Gly378) which lies at the amino terminus of the active site helix, Cα3.

A peculiar feature of the active site is its position with respect to the two central α3-helices (Figs 9 and 10). These two helices point with their amino termini towards the active site, suggesting that the helical dipole [18] might assist in catalysis. However, none of the three catalytic residues is positioned on the helical axis. The loops from β3, leading into helix α3, contain the catalytic residues, Cys125 in the N-domain and His375 in the C-domain. This loop is only four residues long in the N-domain, but comprises 14 residues in the C-domain. Both loops have a residue with a positive φ; residue Gln124 in the N-domain (φ,ψ=73°,−132°) and residue Leu377 in the C-domain (φ,ψ=7°,−76°). Gln124 and Leu377 are in equivalent positions with respect to the overall fold of the N- and C-domains, as well as with respect to the beginning of helix α3 (Fig. 2). The structures of these residues are well defined by the electron density map, therefore it seems that some strain in the folded protein is required at these positions in order to achieve proper positioning of the active site residues [19].

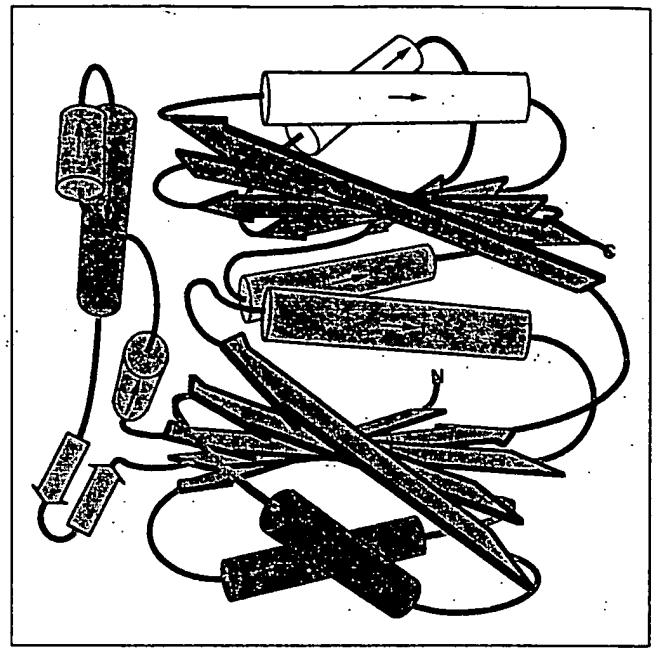

Fig. 6. Schematic illustration of the five-layered structure of thiolase. The active site is locatde near the amino-terminal ends of the green helices of the center layer. The secondary structure elements of the loop domain are in lilac. (Figure prepared by JPh Zeelen.)

Table 2. Active site residues.

Domain	Region	Residues[a]
N	Loop Nβ2 to Nα2	Val93
N	Loop Nβ3 to Nα3	Arg123 Gln124 **Cys125** Ser126 Ser127
L	Loop before Lα2	**Met186** Thr189
C	Loop Cβ1 to Cα1	**Met315** Gly316 Pro319
C	Loop Cβ2 to Cα2	**Asn343** Glu344 **Ala345** Phe346 Gln349
C	Loop Cβ3 to Cα3	Ala370 **His375** Pro376 Leu377 Gly378 Thr380 Gly381 Gln384
C	Loop Cβ4 to Cβ5	Ser401 Met402 **Cys403 Ile404 Gly405** Thr406 Gly407 Met408 Gly409

[a]All residues with atoms within 9 Å of the center of the active site of subunit-1 are tabulated. The bold residues have atoms within 5 Å of the active site center.

Subunit–subunit interactions

The two subunits make numerous interactions including hydrophobic contacts, hydrogen bonds, and salt bridges. There are 239 atom pairs within a cutoff distance of 4 Å. As shown in Fig. 11, three layers of interactions can be recognized. In the first layer, on the side of the molecule

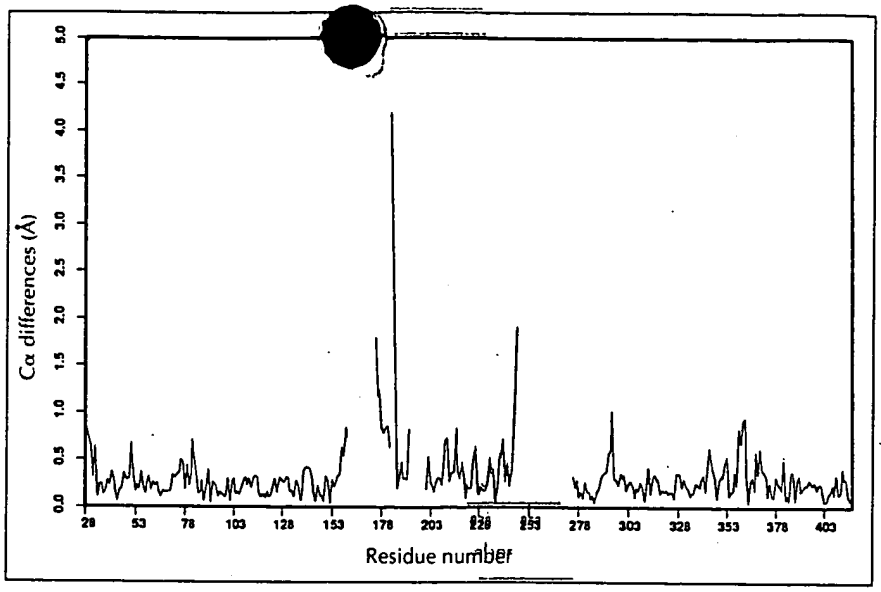

Fig. 7. The Cα-differences between the two subunits. The Cα–Cα distance (in Å) between equivalent Cα atoms of subunit-1 and subunit-2 is plotted on the vertical axis against the residue number. The largest differences are in the loop domain (residues 155 to 276).

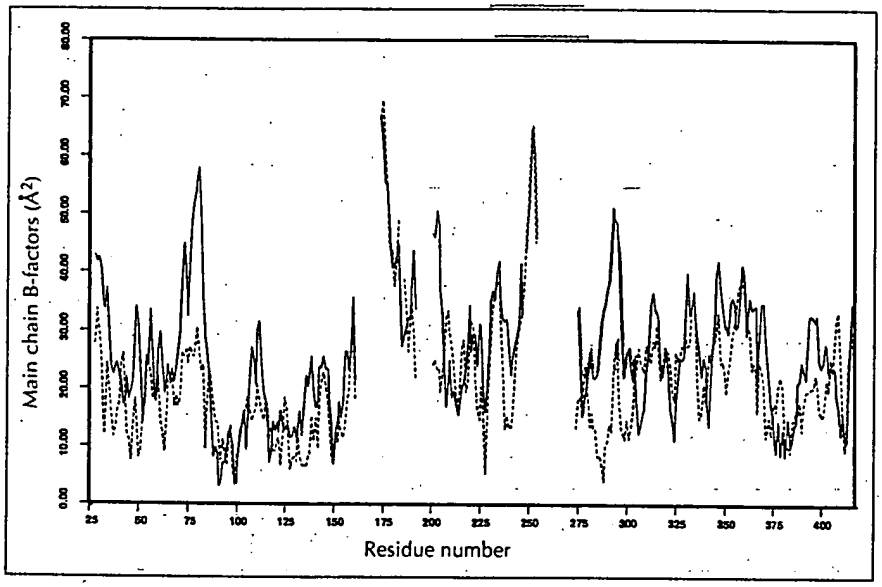

Fig. 8. The average main chain (N,Cα,C,O) B-factors (Å²) of subunit-1 (continuous line) and subunit-2 (discontinuous line) plotted against residue number.

where the two active site pockets are located (Fig. 3), two loops approach and contact each other (Fig. 9). These loops are between Nβ2 and Nα2. In particular, residues 95 to 98, which form a type II β-turn, interact with each other through hydrogen bonds between main chain atoms. The helix Nα2 also interacts with residues from the loop domain of the other subunit (Fig. 9). For example, Cβ(Ala179) and Sγ(Cys182) of the loop domain fit into a pocket formed by hydrophobic atoms of the side chains of Arg104, Ala105, Leu108 and Ala109 of Nα2 of the other subunit. Most of the interactions are seen in the second layer. In this layer, the Nβ3 strands of subunit-1 and subunit-2 form an antiparallel pair of β-strands, related by the dimer two-fold axis (Fig. 9). This interaction means that the mixed β-sheets of the two subunits are continuous, forming a 10-stranded mixed β-sheet. The third layer of interactions is between side chains at the carboxyl termini of the two active site Nα3 helices (Fig. 11). The charged side chains of the Nα3 residues Asp134 and Lys138 point to a polar pocket between the subunits. A salt bridge network is observed from Oδ1(Asp134) to Nζ(Lys138) to Oϵ2(Glu87) (within the same subunit) to Nη2(Arg123) of the other subunit. From this analysis it is clear that most of the subunit–subunit interactions are mediated by contacts between atoms of the two N-domains.

Discussion

The thiolase subunit is folded as a five-layered structure, with outermost and innermost layers consisting of α-helices, with β-sheets in between. At least one other enzyme, inositol monophosphatase, also adopts a five-layered αβαβα structure [20]. However, the topology of the fold of inositol monophosphatase is completely different. In thiolase, two core domains of identical topology can be recognized. The core domains each consist of five β-strands and three α-helices (Fig. 4). The loops between the secondary structure elements are of different lengths in the N-domain and the

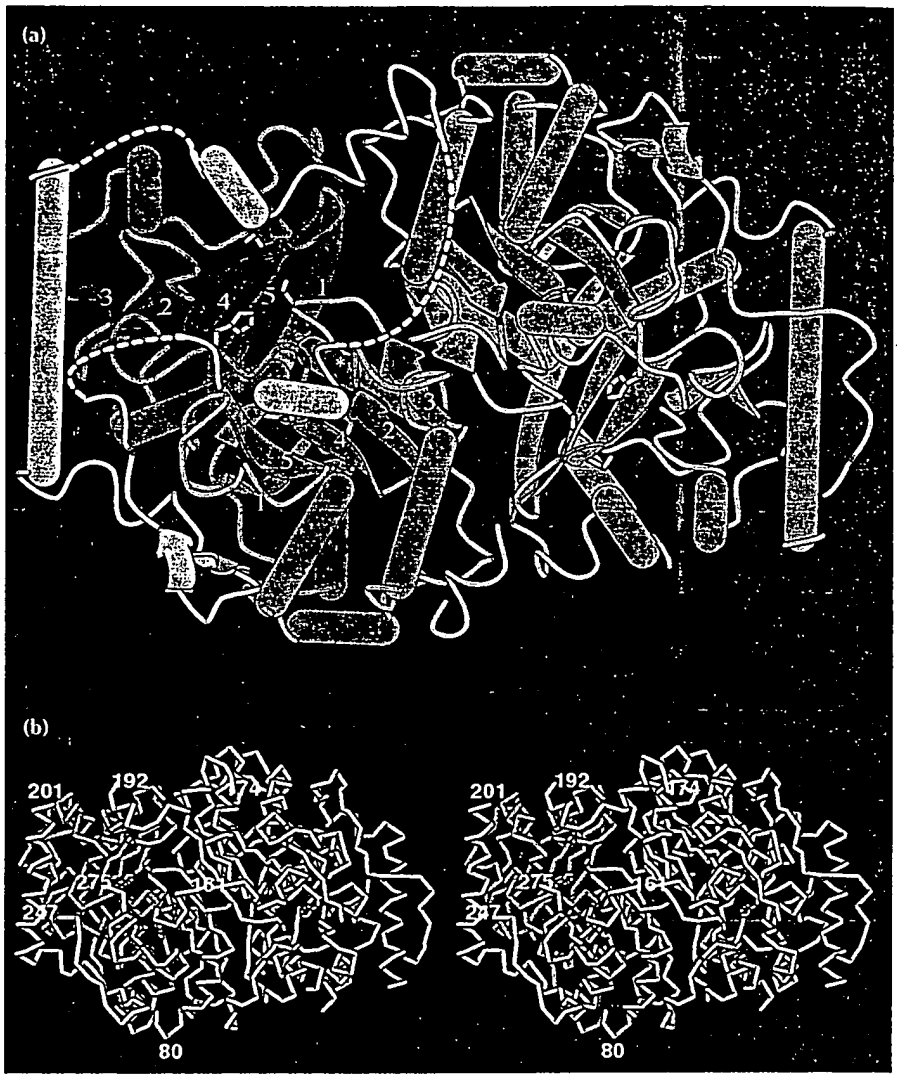

Fig. 9. The thiolase dimer. The view is along the dimer twofold axis, towards the two active sites. The amino and carboxyl termini are on the side of the molecule furthest from the viewer. (a) Schematic drawing of the thiolase dimer with subunit-1 in green (the N-domain), yellow (the loop domain), and red (the C-domain), Subunit-2 is in blue. The dotted yellow lines are the regions of the loop domain which could not be built. The β-strands of the green and red core domains are numbered. The two active site helices, helix Nα3 and Cα3, are marked by asterisks near their amino ends. The side chains of the active site residues [Cys125 (green domain), His375 (red domain) and Cys403 (red domain)] of both active sites are drawn in white. (b) Stereo Cα-trace of the thiolase dimer with the same color scheme as in (a). The residues at the discontinuities of the yellow loop domain are labeled. Other markers along the Cα-trace are position 80 (before Nβ2), Cys125 (after Nβ3), His375 (after Cβ3) and Cys403 (after Cβ4).

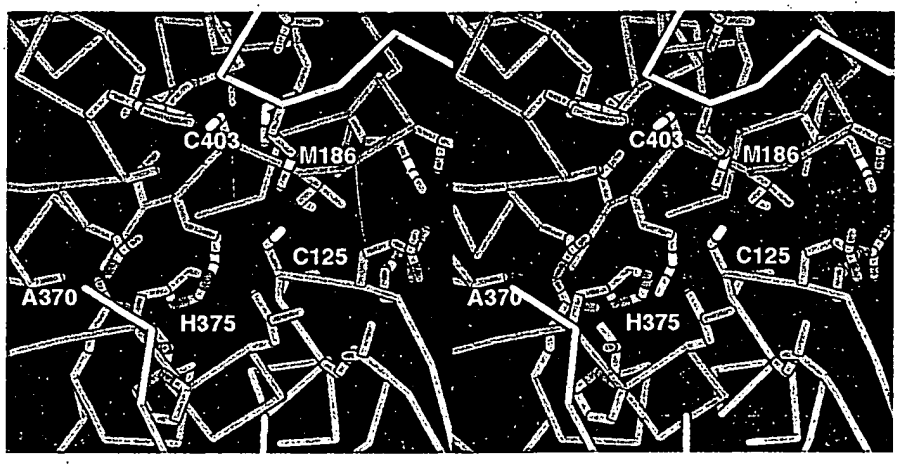

Fig. 10. The active site of subunit-1. The main chain coloring scheme is the same as in Fig. 9. The active site residues Cys125, His375 and Cys403 are labeled. Met186 and Ala370 are also labeled. The displayed side chains are within 9 Å of the center of the active site (Table 2). The active site His375 side chain is surrounded by the side chains of Ala370, Gln343, Cys125 and Leu377. The Cys125 side chain is close to the side chains of His375, Met186, Gln124 and Leu377. The Cys403 side chain is near Gln349 and Met315. Distance information is provided in the text. The two active site helices Nα3 and Cα3 follow Cys125 and His375, respectively.

C-domain. The catalytic residues are found in or near the loops immediately following the β-strands. The topology of the core domain is also present in the first domain of phosphoglucomutase, as detected by the program DALI 21]. According to the structural alignment, shown in Fig. 12, there are 77 equivalent residues (only 7 % sequence identity) which superimpose with an rms difference between Cα positions of 3.1 Å. Phosphoglucomutase is a four-domain enzyme [22] in which domains 2 and 3 have similar topologies to domain 1, but lack the first strand. The assembly of these domains into the complete structure is very different from the thiolase fold. Nevertheless, helix α3 (between β3 and β4) in the first domain of phosphoglucomutase points towards the active site. Also, the turn between β4 and β5 is of particular importance,

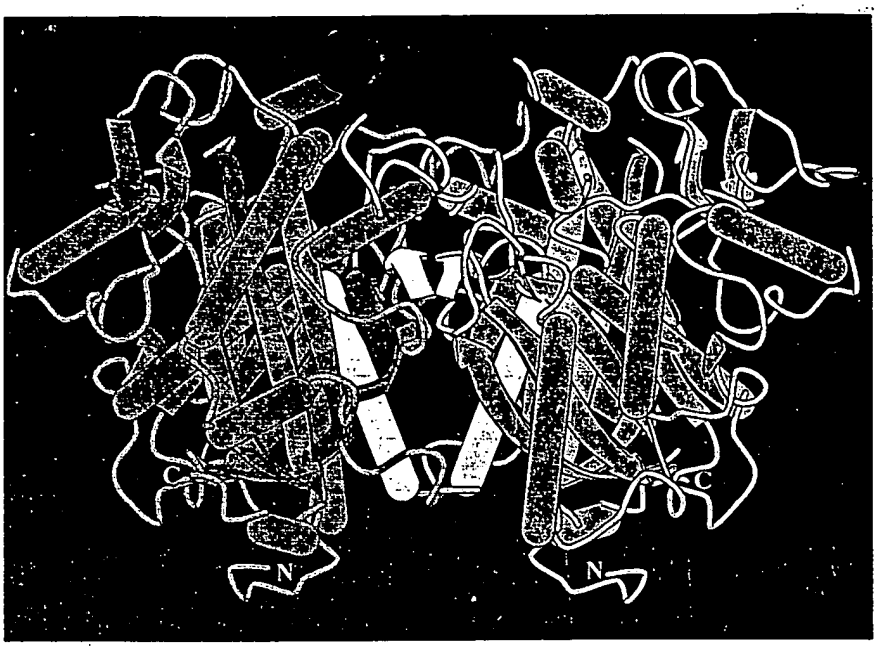

Fig. 11. Side view of the dimer (same view as Fig. 3). The two subunits are shown in pink and blue. The three layers at the dimer interface are highlighted in yellow. The active site face is uppermost, the amino and carboxyl termini (labeled as N and C, respectively) are visible in the lower part of the figure. The amino terminus (Leu28) points into the solvent region. At the top, the yellow loops in contact are between Nβ2 and Nα2. The yellow β-strands of the middle layer are the Nβ3 strands and the dimer interface helices of the bottom layer are the carboxy-terminal ends of the Nα3 active site helices.

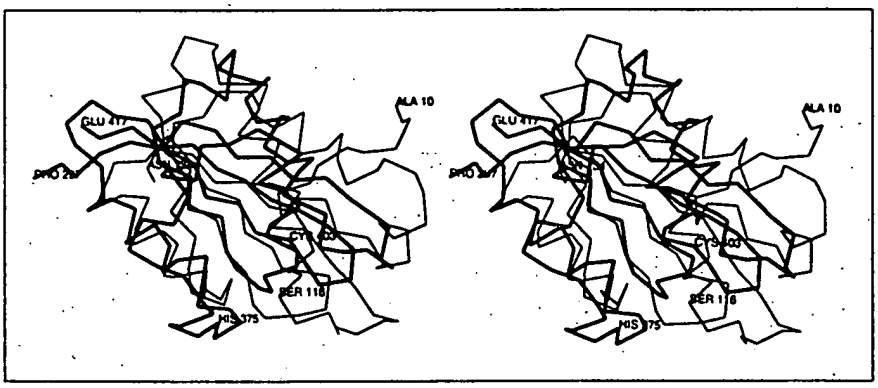

Fig. 12. Stereo superposition of the C-domain of thiolase (thick lines) with the amino-terminal domain of phosphoglucomutase (thin lines). The thiolase domain is labeled at residues Pro297, His375 (active site), Cys403 (active site), and Glu417. The phosphoglucomutase domain [22] is labeled at residues Ala10, Ser116 (active site) and Asn134.

because it contains the active site serine (Ser116). In thiolase, the loop from Nβ3 continuing into Nα3 contains an active site cysteine (Cys125) and the active site histidine, His375, is found in the equivalent loop of the C-domain. The turn between Nβ4 and Nβ5 constitutes the loop domain. This is a short turn in the C-domain and immediately follows the second active site cysteine, Cys403.

The active site of thiolase has been identified and its floor is defined by Cys125, His375 and Cys403. The gold-cyanide ion, $Au(CN)_2^-$, and the gold-chloride ion, $AuCl_4^-$, bind in this pocket between Sγ(Cys403) and Sγ(Cys125). Not surprisingly, gold-chloride is an inhibitor of the thiolase reaction, showing 50 % inhibition at micromolar concentrations. From the covalent modification studies of pig heart thiolase II [4] and *Z. ramigera* thiolase [23] it can be concluded that Cys125 of yeast thiolase will be acylated after the first half reaction (Fig. 13) has been completed. Further mechanistic studies with *Z. ramigera* thiolase [9] indicate that Cys403, and probably also His375, are important catalytic residues. The crystal structure of yeast thiolase confirms that these residues are in the active site pocket and that the two cysteines are within 14 Å of each other, in agreement with spectroscopic measurements [8]. The active site pocket is a small surface patch, shaped by residues of the two core domains and surrounded by the loop domain, all from the same subunit (Fig. 9). The substrate, 3-ketoacyl-CoA, is a large molecule. Therefore, it seems likely that some residues of the loop domain are important for binding the substrate. Evidence for this also comes from the observation that 11 residues of the loop domain are conserved in 21 thiolase sequences (Fig. 2). Two of these residues, Gly266 and Ala270, are in the disordered part of the loop domain. Since these residues are conserved in sequences of both thiolase I and thiolase II they are probably important for the binding of the CoA moiety, which is common to the substrates of both enzymes, rather than for the binding of the fatty acid moiety, which is only present in the substrate of thiolase I.

Although the two active site helices form the central layer, these helices are not particularly hydrophobic (Fig. 2). A buried arginine (Arg383) is observed near the amino terminus of Cα3, which forms a salt bridge with Glu341. Arg383 is a conserved residue (Fig. 2), whereas Glu341 is either a glutamate or an aspartate in all sequences. Nα3 has charged residues at its carboxyl

terminus. These residues face a polar pocket between the two subunits (Fig. 11) and are involved in a salt bridge network across the dimer interface.

Comparison of the sequence alignment of the seven peroxisomal 3-ketoacyl-CoA thiolases with the sequence alignment of 21 thiolases deposited in databanks (data not shown) shows that the seven peroxisomal thiolases are a much closer family (152 identical residues out of 417) compared with the complete set (40 identical residues out of 417). These 40 completely conserved residues are circled in Fig. 2 and many are near the active site.

Another common feature of the seven peroxisomal sequences is the amino-terminal extension of 30 to 50 residues before the beginning of the first β-strand (Nβ1), while only a few more residues are present after the last β-strand (Cβ5). As is shown in Figs 3 and 11, the amino and carboxyl termini are close together and exposed to the solvent. Leu28 is the first residue at the amino terminus that is visible in the electron density map. The observation that the protein in the thiolase crystals is a mixture of the complete primary translation product, and polypeptides starting with amino acid residues 5 and 7 is interesting because this thiolase has been shown to possess an amino-terminal peroxisomal targeting signal (PTS2). The occurrence of a mixture of amino termini probably results from the massive over-expression (about 100-fold) of the thiolase from a multicopy vector. Immunoelectron microscopy revealed that thiolase was present not only in peroxisomes but throughout the cells with the exception of the mitochondria (M Veenhuis and W-H Kunau, unpublished data). If thiolase undergoes processing upon import into peroxisomes, leading to a mature form which lacks the six amino-terminal amino acids, this would explain why a major fraction of the enzyme starts with amino acid Ser7. However, it cannot be excluded that cleavage of the first six amino acids may have resulted from limited proteolysis during protein purification.

It has been demonstrated that the first 16 amino acid residues of thiolase are essential and sufficient for import into yeast peroxisomes [15,16]. The fact that the first 27 residues of the thiolase structure are disordered and exposed to the solvent should allow the interaction of the targeting signal with its putative receptor even in the folded state of the enzyme. On the basis of genetic evidence, a possible candidate for the PTS2 receptor is the product of the *PAS7* gene [24]. Recent data suggest that the Pas7 protein is cytosolic and only binds to peroxisomes in the presence of thiolase (M Marzioch, R Erdmann, M Veenhuis and W-H Kunau, unpublished data). Thus, it now seems important to investigate whether the Pas7 protein binds thiolase in a PTS2-dependent manner and whether such an interaction requires folded or unfolded thiolase.

Genetic diseases have been found which are associated with defects in the genes for human peroxisomal

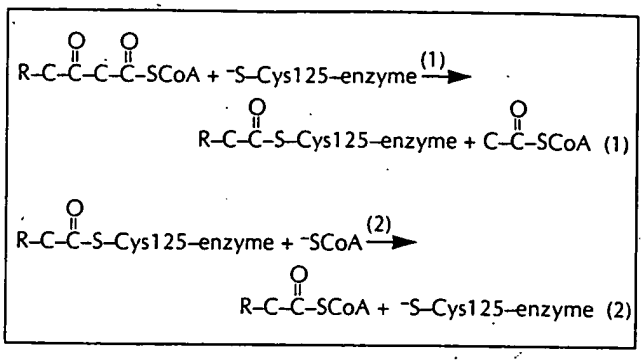

Fig. 13. The two half reactions of thiolase I (adapted from [9]). In reaction (1), acetyl-CoA is split off from a ketoacyl-CoA molecule and a covalent acyl–enzyme intermediate is formed. In reaction (2), CoA reacts with the covalent intermediate and an acyl-CoA molecule is released.

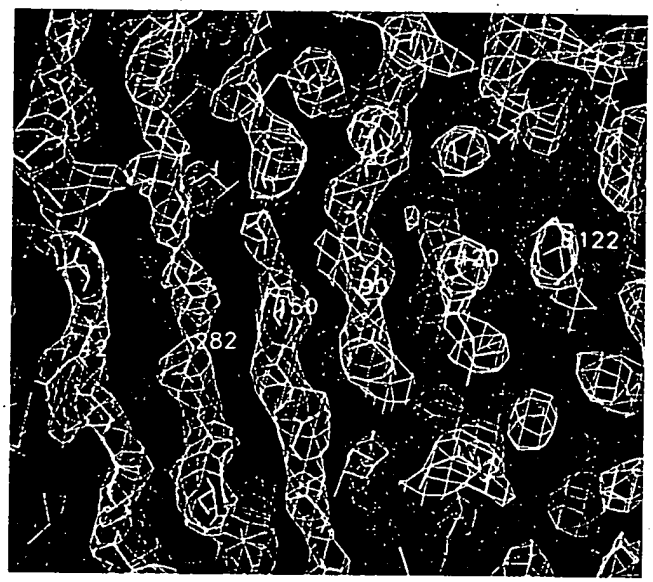

Fig. 14. The final averaged and solvent-flattened 3.1 Å MIRAS map, contoured at 1.2σ. Residues 42, 282, 150, 90, 120 are in the β-strands of the N-domain of subunit-1 and residue B122 is in β-strand Nβ3 of subunit-2.

3-ketoacyl-CoA thiolase [25] and for human mitochondrial acetoacetyl-CoA thiolase [26]. In the latter case, these genetic defects, which cause severe health problems, have been attributed to point mutations. In one patient, a Gly→Arg mutation was found at a position equivalent to residue Asn176 [27]. This position is reasonably well conserved in the non-peroxisomal thiolases. In another patient, an Ala→Thr mutation was found [28]. The equivalent residue in yeast thiolase is Ala370, which is close to the active site His375 (Fig. 10), and conserved in all thiolases (Fig. 2). Experimentally, it is found that both the Gly→Arg mutation and the Ala→Thr mutation result in thiolase variants which are less stable than wild-type thiolase. From the structure it can be seen that the Ala370→Thr mutation could influence the allowed side chain conformation of His375, and therefore interfere with the proper functioning of the thiolase. The consequences of the

Gly176→Arg mutation are trivial to explain from the structure of the dimeric yeast thiolase, because it is a surface residue (at the beginning of the loop domain). Since this point mutation is observed in a tetrameric thiolase, it is possible that this region of the molecule is buried at one of the additional protein interfaces of the tetramer. However, the quaternary structure of the tetramers is unknown. The structure of the yeast thiolase dimer, described in this paper, might provide some clues as to how the dimers could assemble into tetramers.

Biological implications

Thiolases play an essential role in the biodegradation of fatty acids in the β-oxidation pathway, as well as in biosynthetic pathways, where they are important for the Claisen condensation reaction between two molecules of acetyl-coenzyme A (acetyl-CoA) to form acetoacetyl-CoA, which is subsequently used as a building block for further synthesis. The biosynthetic and biodegradative thiolases have similar sequences, and use the same reaction mechanism.

The crystal structure of the biodegradative 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae* peroxisomes provides the first description of the three-dimensional architecture of this class of enzymes. This thiolase is dimeric and is involved in the β-oxidation pathway. The extensive interactions that are observed across the dimer interface suggest that it is a very stable dimer.

The thiolase dimer has one extensive and rather flat side, perpendicular to the dimer two-fold axis. Both active sites are located on this side of the molecule, near the dimer interface, whereas the amino and carboxyl termini of both subunits are found on the opposite side. A mutation in the thiolase gene causing severe health problems, such as ketoacidotic attacks at the age of six months followed by severe retardation, has been mapped to the active site region and the structure provides an explanation for the detrimental effects of this mutation. The amino-terminal 27 residues are present in the crystallized protein, but are not visible in the electron density map. These residues also contain the peroxisomal targeting sequence. The crystal structure shows that this targeting sequence is disordered in the folded thiolase structure.

In the unliganded structure decribed here, the active site is a shallow pocket surrounded by residues of the loop domain. Some residues of this loop domain are disordered, so the exact shape of the binding pocket is unknown. Some of these disordered residues are completely conserved in 21 thiolase sequences, suggesting that these loop domain residues are important for the formation of the enzyme–ligand complex which then must acquire its specific structure by an induced fit mechanism. The crystal structure shows that the side chains of three conserved catalytic residues, Cys125, His375 and Cys403 form the base of the active site pocket. The special properties of Sγ(Cys125) and Sγ(Cys403) in the two-step reaction must be related to their interactions with neighboring main chain and side chain atoms. More mechanistic and structural binding studies will be required to elucidate further details of the reaction mechanism.

Materials and methods

Expression and purification

In order to facilitate a large-scale purification of thiolase (Fox3p) we constructed a thiolase overproducing *S. cerevisiae* strain, AY-FOX3, in the following way: a 2.5 kb genomic *Sau*3A fragment, functionally complementing thiolase-deficient mutants of *S. cerevisiae* (R. Erdmann & W-H Kunau, unpublished data), was subcloned into the *Bcl*I-digested pCS19 vector [29]. This subcloning introduced a *Xho*I site at the 5′ end (about 900 bp upstream of the *Bcl*I site) and a *Bam*HI site at the 3′ end (about 600 bp downstream of the *Bcl*I site) of the gene. Subsequent *Xho*I/*Bam*HI digestion resulted in a 4.0 kb fragment which was inserted into a *Bam*HI/*Sal*I cut multicopy vector, YEp352 [30], to form YEp352-FOX3. Transformation of the protease-deficient *S. cerevisiae* strain ABYSδ6 (a gift from D Wolf, Stuttgart, Germany) with plasmid YEp352-FOX3 yielded the strain AY-FOX3. Growth on oleate as the carbon source [31] resulted in a 70-fold to 90-fold increase in thiolase activity over wild-type levels. Thiolase was purified from 8–10 g of oleate-grown cells (wet weight) according to a published procedure [31].

Crystallization, data collection and heavy atom derivatives

The crystals were grown using a microdialysis method [32,33]. The protein solution [5 mg ml^{-1} in a buffer of 200 mM potassium phosphate pH 7.4, 1 mM dithiothreitol (DTT), 1 mM NaN_3, 1 mM EDTA] was dialyzed against 1 ml of a buffer of 25 mM 3-(*N*-morpholino)-propanesulfonic acid (MOPS) pH 6.5, 1 mM DTT, NaN_3, EDTA. The space group of the crystals is $P2_12_12_1$ with cell dimensions of 71.78 Å, 93.72 Å, and 120.45 Å and with one dimer per asymmetric unit.

The crystals grow at room temperature in about one week. Mass spectrometry measurements using protein from dissolved crystals show that the crystallized yeast thiolase consists of a mixture of proteins with molecular weights varying from 44.6 kDa to 44.0 kDa. This is in good agreement with the amino-terminal amino acid sequence analysis of dissolved crystals, which also showed a mixture of fragments, starting at positions 2, 5 and 7 with predominant occurrence of the sequence SIKD, which corresponds to residues 7 to 10. Clearly, the crystallized protein includes all residues after Ser7. The complete protein, starting at Ser2 is only present with an estimated occupancy of less than 50 %.

In order to prevent deterioration of crystal quality the DTT concentration of the dialysis buffer was raised to 2 mM one week after setting up the crystallization experiment. Every

month, the dialysis solution was refreshed in order to be sure of a high concentration of DTT in its reduced state. Three days before data collection, the DTT concentration was raised to 200 mM. A complete native dataset was collected from one crystal with a maximum resolution of 2.8 Å (Table 3). The procedure resulting in the first interpretable MIRAS map has been described elsewhere [34]. Most of the calculations were done with CCP4 programs [SERC (UK) Collaborative Computing Project 4, Daresbury Laboratory, UK, 1979]. Briefly, poor MIRAS phases at 3.1 Å resolution, with <m>=0.47, were calculated from six heavy atom derivatives (Table 4). The best heavy atom derivative was gold-chloride,

Table 3. Statistics of the native dataset.

Total observed reflections (resolution 30 Å–2.8 Å)	47 231
Total unique reflections (resolution 30 Å–2.8 Å)	18 981
Merging R-factor[a]	7.9%
Overall completeness (resolution 30 Å–2.8 Å)	92.0%
Overall completeness (resolution 2.9 Å–2.8 Å)	92.6%
Space group	$P2_12_12_1$
Cell dimensions (Å)	71.78, 93.72, 120.45

[a] $R_{merge} = \Sigma_h\Sigma_i|< I >_h - I_{h,i}| / \Sigma_h\Sigma_i|< I >_h|$

Table 4. Heavy atom data.

	Maximum resolution	R_{merge}[a]	Number of sites	R_{cullis}[b] (centric data)	Phasing power[c] centric/acentric
$KAuCl_4$ (1)	2.71 Å	6.8 %	2	0.75	1.2/0.9
$KAuCl_4$ (2)	3.02 Å	8.4 %	2	0.81	0.8/0.5
$KAu(CN)_2$	2.71 Å	7.7 %	2	0.80	0.8/0.6
$KAuI_4$	3.23 Å	9.1 %	2	0.97	0.6/0.5
$K_3UO_2F_5$	4.02 Å	12.4 %	2	0.86 (to 4 Å)	0.7/0.5 (to 4 Å)
$(NH_4)_2U_2O_7$	3.10 Å	8.9 %	3	0.94	0.5/0.4

[a] $R_{merge} = \Sigma_h\Sigma_i|< I >_h - I_{h,i}|/\Sigma_h\Sigma_i|< I >_h|$ [b] $R_{cullis} = \Sigma||F_{PH} \pm F_P| - F_H|/\Sigma|F_{PH} \pm F_P|$ for centric reflections. [c] Phasing power = $rms(F_H)/rms(E)$ where F_P, F_{PH}, F_H are the native, derivative and heavy atom structure factor amplitudes respectively, and E the lack of closure.

with two heavy atom sites per dimer. The orientation and position of the local dimer two-fold axis were determined with the GLRF-package [35,36], using the self rotation function and translation function options, respectively.

The MIRAS map was much improved by two-fold averaging using program O [37]. A partial model was built in the averaged map. A complete model could eventually be built in a DEMON-averaged map [38], after improving the mask, optimizing the local transformation and changing the weighting scheme [34]. After trying to refine this structure, it appeared that parts of the loop domain were not related by the dimer two-fold axis. These regions were omitted from the model and another (dual-)mask was generated, allowing for; averaging of the map in the regions of the core domains; no averaging in the regions of the loop domain, and solvent flattening in the regions outside the protein. This mask (and the associated averaged MIRAS map) was further improved iteratively in the course of subsequent refinement. The averaged and solvent-flattened 3.1 Å MIRAS map calculated with the final mask was used as a reference map throughout the final model building sessions and refinement calculations. This map is of good quality. Fig. 14 shows the quality of this map near the β-sheet of the N-domain of subunit-1.

Structure refinement and structure analysis

The structure was refined in an iterative procedure, using, alternately, X-PLOR calculations [39] and model building into electron density maps. The X-PLOR refinement was done first at 3.1 Å with strict non-crystallographic symmetry constraints, and then at 2.8 Å with non-crystallographic symmetry restraints. The X-PLOR models were compared with SIGMAA weighted $2F_o-F_c$ maps [40] in extensive model building sessions, using the O package [41]. Parts of the loop domain which had been omitted could be rebuilt in the electron density map as the refinement progressed. Some regions of the loop domain are well defined, for example the long helix Lα3 and the two β-strands Lβ1 and Lβ2. The model building in $2F_o-F_c$ maps was guided by positive and negative peaks in the corresponding F_o-F_c maps and by the averaged and solvent-flattened 3.1 Å MIRAS reference map. In the final model, there are four fragments missing in subunit-1, and five fragments in subunit-2, owing to weak or discontinuous density in the corresponding sections of the electron density map.

The final model has an R-factor of 19.8 %, with a free R-factor of 33.4 % (Table 1). Water molecules have not been included in the refinement at this stage. The complete model consists of 5201 protein atoms. The missing residues of subunit-1 are 1–27, 162–173, 193–200, 248–274 and residues 1–27, 162–174, 184–185, 193–200, and 255–273 of subunit-2 are not included in the model. The Ramachandran plot of the remaining residues shows that there are 11 residues which are not close to the allowed regions: five in subunit-1 (50, 82, 124, 377 and 395) and six in subunit-2 (50, 124, 177, 178, 182, 251 and 377). These residues have higher than average main chain B-factors, except for residues 124 and 377 in subunit-1, and residues 50, 124, 377 in subunit-2. The average B-factors, calculated from all atoms, for the N-domains, loop domains and C-domains are 21 Å^2, 33 Å^2 and 25 Å^2, respectively. The N-domains have the lowest average B-factor, which correlates with their position in the center of the dimer, near the dimer two-fold axis (Fig. 9). The average B-factors of subunit-1 and subunit-2 are 28 Å^2 and 22 Å^2, respectively. For the core domains there are only three significant chain breaks at the 1.2σ level; they occur at positions 151 (in Nβ4 of subunit-1), 145 (between Nα3 and Nβ4 of subunit-2) and 335 (between Cα1 and Cβ2 of subunit-2). As shown in Table 1, the geometry of the structure is in good agreement with ideal geometry. The two gold sites are in the two active site pockets, at a distance of 3.1 Å from Sγ(Cys125) and 2.4 Å from Sγ(Cys403) in subunit-1 and 3.2 Å from Sγ(Cys125) and 2.5 Å from Sγ(Cys403) in subunit-2.

Subunit-1 has been used as the reference subunit for the structure analysis. The package O [41] and WHAT IF [42] have been used for structure analysis calculations. Figs 9, 10 and 11 have been drawn by XOBJECTS (MEM Noble, Oxford University, unpublished program). The sequence alignments used for Fig. 2 were calculated by the PILEUP program of the GCG-package [43]. The sequences were taken from sequence databanks.

The coordinates have been deposited in the Brookhaven Protein Data Bank (accession code 1PXT).

Acknowledgements: We thank Ursu[redacted]pmund for the purification of thiolase, Dr Read (Edmonton) for help with the averaging protocol, Dr Holm (EMBL) for carrying out the database search with the program DALI, and Prof Petsko (Boston) for pointing out the importance of high concentrations of DTT to prevent deterioration of crystal quality. The sequence and mass spectrometric analysis of thiolase were done by the EMBL peptide group.

References

1. Middleton, B. (1973). The oxoacyl-coenzyme A thiolases of animal tissues. *Biochem. J.* **132**, 717–730.
2. Kunau, W.-H., Dommes, Y. & Schulz, H. (1994). β-Oxidation of fatty acids in mitochondria, peroxisomes and bacteria. A century of continued progress. *Prog. Lipid Res.*, in press.
3. Masamune, S., Walsh, C.T., Sinskey, A.J. & Peoples, O.P. (1989). Poly-(R)-3-hydroxybutyrate (PHB) biosynthesis: mechanistic studies on the biological Claisen condensation catalyzed by β-ketoacyl thiolase. *Pure & Appl. Chem.* **61**, 303–312.
4. Gehring, U. & Harris, J.I. (1970). The active site cysteines of thiolase. *Eur. J. Biochem.* **16**, 492–498.
5. Staack, H., Binstock, J.F. & Schulz, H. (1978). Purification and properties of a pig heart thiolase with broad chain length specificity and comparison of thiolases from pig heart and *Escherichia coli*. *J. Biol. Chem.* **253**, 1827–1831.
6. Izbicka-Dimitrijevio, E. & Gilbert, H.F. (1982). Two sulfhydryl groups near the active site of thiolase I from porcine heart: modification of thiolase with the fluorescent thiol reagent S-mercurio-*N*-dansyl-L-cysteine. *Biochemistry* **21**, 6112–6118.
7. Salam, W.H. & Bloxham, D.P. (1986). Identification of subsidiary catalytic groups at the active site of β-ketoacyl-CoA thiolase by covalent modification of the protein. *Biochim. Biophys. Acta* **873**, 321–330.
8. Izbicka, E. & Gilbert, H.F. (1984). Fluorescence energy transfer measurements of spatial relationships between sulfhydryl groups of thiolase I from porcine heart. *Biochemistry* **23**, 6383–6388.
9. Williams, S.F., Palmer, M.A.J., Peoples, O.P., Walsh, C.T., Sinskey, A.J. & Masamune, S. (1992). Biosynthetic thiolase from *Zoogloea ramigera*. Mutagenesis of the putative active-site base Cys378 changes the partitioning of the acetyl S-enzyme intermediate. *J. Biol. Chem.* **267**, 16041–16043.
10. Igual, J.C., Gonzalez-Bosch, C., Dopazo, J. & Pérez-Ortin, J.E. (1992). Phylogenetic analysis of the thiolase family. Implications for the evolutionary origin of peroxisomes. *J. Mol. Evol.* **35**, 147–155.
11. Kurihara, T., Ueda, M. & Tanaka, A. (1989). Peroxisomal acetoacetyl-CoA thiolase and 3-ketoacyl-CoA thiolase from an n-alkane-utilizing yeast, *Candida tropicalis*: purification and characterization. *J. Biochem.* **106**, 474–478.
12. Gehring, U. & Harris, J.I. (1970). The subunit structure of thiolase. *Eur. J. Biochem.* **16**, 487–491.
13. Swinkels, B.W., Gould, S.J., Bodnar, A.G., Rachubinski, R.A. & Subramani, S. (1991). A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-ketoacyl-CoA thiolase. *EMBO J.* **10**, 3255–3262.
14. Osumi, T., *et al.*, & Hashimoto, T. (1991). Amino-terminal presequence of the precursor of peroxisomal 3-ketoacyl-CoA thiolase is a cleavable signal peptide for peroxisomal targeting. *Biochem. Biophys. Res. Commun.* **181**, 947–954.
15. Erdmann, R. (1994). The peroxisomal targeting signal of 3-oxoacyl-CoA thiolase from *Saccharomyces cerevisiae*. *Yeast* **10**, 935–944.
16. Glover, J.R., Andrews, D.W., Subramani, S. & Rachubinski, R.A. (1994). Mutagenesis of the amino targeting signal of *Saccharomyces cerevisiae* 3-ketoacyl-CoA thiolase reveals conserved amino acids required for import into peroxisomes *in vivo*. *J. Biol. Chem.* **269**, 7558–7563.
17. Richardson, J.S. (1977). β-sheet topology and the relatedness of proteins. *Nature* **268**, 495–500.
18. Hol, W.G.J. (1985). The role of the α-helix dipole in protein function and structure. *Prog. Biophys. Molec. Biol.* **45**, 149–195.
19. Herzberg, O. & Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* **11**, 223–239.
20. Bone, R., Springer, J.P. & Atack, J.R. (1992). Structure of inositol monophosphatase, the putative target of lithium therapy. *Proc. Natl. Acad. Sci. U S A* **89**, 10031–10035.
21. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
22. Dai, J.-B., Liu, Y., Ray, W.J.J. & Konno, M. (1992). The crystal structure of muscle phosphoglucomutase refined at 2.7-angstrom resolution. *J. Biol. Chem.* **267**, 6322–6337.
23. Thompson, S., Mayer[redacted]eoples, O.P., Masamune, S., Sinskey, A.J. & Walsh, C.T. (19[redacted]). Mechanistic studies on β-ketoacyl thiolase from *Zoogloea ramigera*: identification of the active-site nucleophile as Cys89, its mutation to Ser89, and kinetic and thermodynamic characterization of wild-type and mutant enzymes. *Biochemistry* **28**, 5735–5742.
24. Kunau, W.-H., *et al.*, & Wiebel, F.F. (1993). Two complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: forward and reversed genetics. *Biochemie* **75**, 209–224.
25. Schram, A.W., *et al.*, & Wanders, R.J.A. (1987). Human peroxisomal 3-oxoacyl-coenzyme A thiolase deficiency. *Proc. Natl. Acad. Sci. U S A* **84**, 2494–2496.
26. Kano, M., Fukao, T., Yamaguchi, S., Orii, T., Osumi, T. & Hashimoto, T. (1991). Structure and expression of the human mitochondrial acetoacetyl-CoA thiolase-encoding gene. *Gene* **109**, 285–290.
27. Fukao, T., Yamaguchi, S., Orii, T., Schutgens, R.B.H., Osumi, T. & Hashimoto, T. (1992). Identification of three mutant alleles of the gene for mitochondrial acetoacetyl-coenzyme A thiolase. A complete analysis of two generations of a family with 3-ketothiolase deficiency. *J. Clin. Invest.* **89**, 474–479.
28. Fukao, T., *et al.*, & Hashimoto, T. (1991). Evidence for a structural mutation (Ala347 to Thr) in a German family with 3-ketothiolase deficiency. *Biochem. Biophys. Res. Commun.* **179**, 124–129.
29. Sengstag, C. & Hinnen, A. (1987). The sequence of the *Saccharomyces cerevisiae* gene *PHO2* codes for a regulatory protein with unusual aminoacid composition. *Nucleic Acids Res.* **15**, 233–246.
30. Hill, J.E., Myers, A.M., Koerner, T.J. & Tzagoloff, A. (1986). Yeast/*E. coli* shuttle vectors with multiple unique restriction sites. *Yeast* **2**, 163–167.
31. Erdmann, R. & Kunau, W.-H. (1994). Purification and immunolocalization of the peroxisomal 3-oxoacyl-CoA thiolase of *Saccharomyces cerevisiae*. *Yeast*, in press.
32. Zeelen, J.Ph., Wierenga, R.K., Erdmann, R. & Kunau, W.-H. (1990). Crystallographic studies of 3-ketoacylCoA thiolase from yeast *Saccharomyces cerevisiae*. *J. Mol. Biol.* **215**, 211–213.
33. Zeelen, J.Ph. & Wierenga, R.K. (1992). The growth of yeast thiolase crystals using a polyacrylamide gel as dialysis membrane. *J. Crystal Growth* **122**, 194–198.
34. Mathieu, M. & Wierenga, R.K. (1994), From a poor MIRAS-map to a final structure: importance of 2-fold averaging in the structure determination of thiolase. In *Proceedings of CCP4 Study Weekend.* (Waller, D. and Hubbard, R., eds), Daresbury Laboratory, UK, in press.
35. Tong, L. & Rossmann, M.G. (1990). The locked rotation function. *Acta Crystallogr. A* **46**, 783–792.
36. Tong, L., Choi, H.-K., Minor, W. & Rossmann, M.G. (1992). The structure determination of Sindbis virus core protein using isomorphous replacement and molecular replacement averaging between two crystal forms. *Acta Crystallogr. A* **48**, 430–442.
37. Jones, T.A. (1992). A set of averaging programs. In *Proceedings of the CCP4 Study Weekend.* (Dodson, E.J. Gover, S. and Wolf, W. eds), pp. 91–105, Daresbury Laboratory, UK.
38. Vellieux, F.M.D., *et al.*, & Hol, W.G.J. (1993). Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. *Proc. Natl. Acad. Sci. USA* **90**, 2355–2359.
39. Brünger, A.T. (1992). *X-PLOR: Version 3.1: A System for X-ray Crystallography and NMR.* Yale University, New Haven, CT.
40. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149.
41. Jones, T.A., Zou, J.-Y. & Cowan, S.W. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
42. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56.
43. Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.
44. Noble, M.E.M., Zeelen, J.P. & Wierenga, R.K. (1993). Structure of triosephosphate isomerase from *Escherichia coli* determined at 2.6 Å resolution. *Acta Crystallogr. D* **49**, 403–417.
45. Masamune, S., *et al.*, & Walsh, C.T. (1989). Bio-Claisen condensation catalyzed by thiolase from *Zoogloea ramigera*. Active site cysteine residues. *J. Am. Chem. Soc.* **111**, 1879–1881.

Received: **27 May 1994**; revisions requested: **23 Jun 1994**; revisions received: **6 Jul 1994**: Accepted: **8 July 1994.**